Fako Berkers (4131 words)

I copy pasted my (temporarily) what, how and why into this document and wrote a short abstract. Apart from that I started to write down my thoughts as clear as possible under "Being tracked". I still have to take out the sharp edges at the end of the argumentation. I think it shows my opinion too much. Right after that part is the rest of my outline followed by loads of annotations from news items that seem relevant.

Abstract

This thesis explores how people are being tracked online and why this may be a problem or how this can be justified. My project is an attempt to make clear which risks and changes are involved concerning the huge amount of data that gets generated around our personas in a playful manner.

What

When people look at the game they will immediately recognize Monopoly. However there is something strange going on. The streets don't represent streets in Atlantis City or any other city in the world. Instead they portray internet platforms like Facebook, Hotmail, Youtube and Twitter. When you buy part of a set you actually buy share from these web services and although you can't buy houses or hotels you're able to buy investments in ad and tracking technology. The "change" and "public fonds" cards tell stories about how social media are leaking data about their users and how they are affected by this.

The game exists in two forms. The first is an actual organic game, which has been completely altered to depict a different capitalist world we are used to play with when playing Monopoly, as described above. Another form is a digital form that will be played online. A possible mix of these forms may exist as a performance where I move the pieces by hand as others watch on a distant. These turns are not taken instantly, but as with the abstract version only occur two or three times a week.

How

For now I limit my description of how to: a table that states how each element in the original Monopoly game will get restyled to fit the new game reality I want to create.

Monopoly	**WWWonopoly**
Change cards	Messages about data leaks
Public fond cards	Messages about data leaks
Landing on electricity and water company	Draw a change or public fond card
Buying/landing on/trading a street	Buying/landing on/trading shares of a platform
Completing streets of a color	Owning a platform
Buying houses	Investing in ad and track technology (or buying servers!)
Buying hotels	Investing in ad and track technology (or buying a data centre!)
Buying/landing on a railway station	Buying/landing on an internet provider
Get salary	Get salary

Go to prison	Go to prison
Pay taxes	Only two things are sure in life: death and ... ;)
Free parking	Free peer-to-peer downloads
Take a mortgage on property	?

A few rules that are nice to add here:

- There is no such thing as the free p2p download jackpot (as the game would last forever)
- Each player begins with two randomly assigned properties for which they must pay (to speed up play)
- A game will last X number of turns

Why

It lies in our nature that we can't assert risks very well. A lot of people like social media, but very few seem to be aware what the risks of them are. By allowing players to take the perspective of a “system agent”, meaning somebody with power over the functioning of the social media system, the players get a bird view of what is happening. From this perspective the players are more likely to get aware of risks than when they are “in” the system with a frog like perspective. It would be good if by playing the game and reading the bits of text and articles that are attached to each game message, the player gets a better sense of how social media function and where the vulnerabilities of the system lie. Any system can get played, but it’s important that users don’t become victims because of that. By informing users in this somewhat playful manner an early warning could prevent harm.

Being tracked

When I say that you’re tracked online I mean that a lot of things that you do when you are using the web is tracked, saved and used by big companies to make money. I will talk about a few ways in which information about your behavior online is valuable and how it finds its way into the hands of who is interested in it. The fact that you are being tracked is problematic when sensitive information is disclosed to parties who were not supposed to have that information. An example of disclosed information is when you want to surprise your partner with a vacation and the surprise is ruined because commercials for the destination you searched for is shown on a lot of online advertisements. Another example is when insurance companies heighten your yearly fee, because you've been searching for AIDS related topics online.

You may think that this kind of disclose doesn’t apply to you or is not so bad and that you don’t have anything to hide. While this is probably true you can ask yourself whether your friends and family are in the same position and whether you may be in need of privacy as you get older for instance and your medical data becomes more valuable to insurance companies. I hope you come to the same conclusion as me and find that we are better off in a society where we don’t have to worry about whether sensitive information gets revealed or not.

I could sum up a set of incentives which would protect your privacy and stop writing this essay. In fact I have made this list here <<<a link here>>> for anybody who doesn’t want to read further than this. However a rule without an explanation of why that rule is important is just another rule and such rules are likely to get broken. That’s why I want to explain you in simple terms why these rules are important and how they affect your privacy. It will take a little more effort to understand the essay than to read the bullet point list of incentives, but you are more likely to remember and follow rules once you understand their reason. Once you know these rules you can not only protect yourself but also people who are in greater need for it and haven’t red this essay themselves.

A website is a collection of webpages. It's best to picture webpages to be some kind of Power Point slides. This metaphor for what a webpage is works well, because like a Power Point slide a webpage can change, for instance when you click somewhere or automatically after some time. A difference between a PowerPoint slide and webpage is that any changes always become visible, while on a webpage this is not necessarily the case. In fact tracking is for a significant part done by changing the webpage invisibly when you are looking at it!

When you visit a webpage on a website you're using a computer program that is called a browser. Examples of browsers are Internet Explorer, Firefox, Chrome, Safari and Opera. As soon as you click on a link somewhere your browser makes connection with a computer which stands in a room like you see below:

Many powerful computers are stored in those things that look like fancy refrigerators on the image. When you see a webpage in your browser it has been send to your browser by one of these computers.

However there is a catch to this that is important for the way that you're being tracked. Your browser receives an entire webpage in parts. The first part, which usually contains all text on the webpage, will also indicate to a browser where it can find other parts. These parts may be images, video or audio that are on the webpage you're visiting. Since these parts are often essential to the look of the webpage your browser will download these parts without checking if they are necessary. The additional media parts could be located on the same computer as the webpage your visiting, but they could also be present on computers which are in a room on the other side of the world as the computer that gives your browser the webpage. A website logo is often retrieved from the same computer as the first part of any webpage which is part of that website. However some images are almost never on the same computer as the webpage, that shows those images. Think for instance about Facebooks “like it” buttons or Youtubes video players, but also banners and commercials. These media are retrieved from Facebook, Google or another media company. When that happens the computers, from these companies, not only provide the requested media, but also record that you have received the media when visiting a particular webpage from a particular website.

Now the question remains how these 3rd party computers (computers other than the computer delivering the main part of the webpage) know that it is you who is visiting the webpage. Lets make clear that in principle you reveal as much about yourself to a 3rd party computer than to the computer who serves you the main part of a webpage.

Whenever you are asking for a part of a webpage your browser tells a few things to the computer that is supposed to deliver that part to you. It depends on the browser what gets told exactly, but usually it will tell which browser you are using and if you are on a Mac, Windows, Android or other kind of computer. It may also reveal your language setting and which webpage you were visiting before. This information combined already tells a lot about you, because only a few people will send exactly the same information. This makes you identifiable. The really revealing information that gets collected however is the IP address in combination with the content of a so called cookie file.

When your browser connects to a computer for a part of a webpage this computer does need to know where to send the part of the webpage to. To inform websites where you want the webpage parts to be delivered your browser specifies your IP address (Internet Protocol Address). This address that basically functions like your post address is given to you by your internet provider. Any part of a webpage that a browser asks for gets delivered at this address where the browser will put them together to form a whole. Sometimes you share your IP address with the people you live with and it can also happen that you get a partially new address every time you turn on your computer (depends on your internet provider). In general the IP address will limit the size of the group of people who may have asked for the webpage significantly and thus comes very close to identifying you.

The identification is further completed by the use of cookies. Cookies are files that are stored by your browser when you are visiting websites. The content of these files is determined by the computers who send you the parts of a website. So the computers not only send the media or other kind of webpage part you're interested in, but also what the computers want your browser to store in a cookie file. Your browser obeys by default. Whenever you ask again for something from the website, your browser will not only request for the webpage parts, but also remember the website computers what was stored in the cookie file. The content of a cookie file might be the products you've been shopping for so far on a certain website for instance. When you visit the paying webpage the content of the cookie file gets transferred to the website. In this case a correct bill may be generated from this information, since it contains all products you shopped for, which is then shown on the webpage you asked for.

So far so good, but what if a website stores something like a social security number inside a cookie? In that case this identifying number gets send over to the website every time you visit a webpage from that site or download a webpage part from that site. This means that if a certain image or other part of a website is used on lets say a million pages, your travel path between those pages gets tracked precisely. Every time you visit a site from the large collection the same, often useless, web page part is collected from the 3rd party computer. Your browser will tell this computer who you are because your identity gets send over like all information stored in a cookie file. The 3rd party computer will be programmed to record everything you do within those million webpages and relate the information to the big number given to you.

The technique described above is deployed with Facebooks “like it” button. Since this button is present on so many webpages these days Facebook is enabled to track people over the internet. People don't even need to have an account with Facebook to get tracked, nor do you need to click the buttons in order for it to work. Their sheer presence is enough to tell Facebook you've been visiting a page. The “like it” button part is separately retrieved from a computer owned by Facebook and not the website you were originally visiting. Since every request for a webpage part that needs to be collected from Facebook will be done with your browser sending the information in your Facebook cookie file back to Facebook, the unique number created by Facebook for everyone tells Facebook exactly who you are.

If you register with Facebook that is a bonus. Your name and everything else you share on Facebook will get related to the identifying number too. That way they'll know a lot more about you then where you've been online. The same kind of disclosure happens when you visit webpages with your mobile phone. The browser on your mobile phone can send additional information about you, for instance your Google account ID when you're using an Android phone.

Now lets get back to the idea that a webpage is sort of like a Power Point slide which can change automatically or through interactions, without the visitor noticing. Whenever a website changes a 3rd party computer can again be involved for this change and track visitors further. A webpage may at any time connect to a 3rd party computer and inform that computer, which actions a visitor has taken. These actions may range from: where a visitor clicks upon, on which part the user is pointing his/her mouse and for how long or how long the visitor is staying in general. This is possible because a webpage is not so much a page, but rather a dynamic collection of different media originating from different sources.

The kind of tracking I've been explaining so far is called web beacons. There are three alternative ways of tracking which are worthwhile to discuss as well. The first type is called analytics and Google Analytics is probably the largest player in this kind of tracking. This service by Google is used by marketers of various companies who want to know which pages of their website get viewed most often, how long visitors are staying and which links they click to leave the page. They use this information to adjust the website and monitor which marketing actions are effective and which aren't. To make Google Analytics work a webdeveloper has to install Google Analytics programming code on each page that needs to be tracked. What Google Analytics does when it is installed is not only save a unique identifier in cookies, but also the time when you requested for the page and which page you requested before. This way graphs about visitors can be generated for marketers by Google, because the information that is necessary to plot such graphs is gathered by the analytics software.

Recently Google has started a new program called Screenwise. It is an opt-in program, meaning that you have to apply for it, and if you do Google will fetch the Google Analytics kind of information from all websites you visit and not only the ones where Google Analytics is installed. In return for opting-in you get 25 dollar. It's possible that in the long run Google will use the algorithms, that they may be devising with the group who opts-in for Screenwise, upon data collected by their Google Analytics program. Their recent change in privacy policy could be a step in this direction. Currently they seem to say that they are not allowed to interpret the data coming from Google Analytics, but by allowing data to be shared among their services they are only one step away from using Google Analytics data in any of their services. Surely Google will try to sell their use of this data as being user friendly as soon as they deem algorithms coming from the Screenwise program a success. However a lot of people are critical about the current widening of allowances for Google so it may be a while before they dare to push it further.

Another way of tracking the behaviour of unknowing computer users is by installing beacons not on the web, but in emails. These beacons are usually images as well, but they can also be special invisible parts of an email. They work the same way as webbeacons in the sense that when the images in the email get viewed they are downloaded by your email program/website from a website. This download is registered by this site and usually your email address gets send along with the request to download the image. The result is that the website is capable of determining who reads their email and who doesn't. This information is used by spammers who will focus their attacks upon people who open the emails. A lot of email programs block this kind of tracking, but they are sometimes still vulnerable to variants of the same principle.

Next I will explain how all this data can be used to earn money. This will give an idea about how your data is used and where you could become wary for. As said above a lot of your information is used through something like Google Analytics to analyze the visitors use of a website. In the future this information may get used in different ways, but it's not at the moment. Another way for your information to be used is that spammers may act upon which emails you open and which you don't as explained as well.

The main reason where web beacon information is used for is targeted advertisements. You see this kind of commercials everywhere online these days. If you've ever visited a shoe shop online for

instance, changes are great that you get a lot of shoe commercials for weeks to come. Through a web beacon it has been noticed that you visited the shop, but didn't buy anything. To persuade you commercials coming from the same 3rd party computers as the web beacons will consist of shoe ads. In that case the web page part consisting of a commercial block will be generated on the fly, especially targeted at you, or at least at the unique number that your browser has sent towards the ad company.

Companies who follow you for this line of business come in three flavors. They all have in common that they get paid by other companies willing to pay for advertisements in the hope that they will have some effect on your buying behavior.

The first kind of company who watches where you are going and creates a profile based on what they now about the content of the webpages you visited. Whenever possible they'll show you ads that in their opinion are related or relevant in relation to the content you've been reading so far. This may be quite far off. If I have a tooth pain and I visited a forum about that I may get commercials about toothbrushes. It has to do with teeth, but it's not the solution I'm searching for probably.

The second kind of company are called retarget companies. These kind of companies are mostly active in and around webshops. They'll monitor your buying behavior and when you don't buy they'll show you commercials from the products you haven't bought yet. Of course it is their hope that this time you'll get persuaded and make a buy online from the company they are working for.

The last variation of company are working like accountants. It would be easy for a retarget company to say that they've been showing all kinds of commercials about products that users didn't buy before, but how does the seller of these products know that this is true? They'd have to track people themselves in order to know that these companies are not conning them in some way. Instead of tracking people themselves they leave it to specialized companies who monitor the monitors by monitoring you.

Advertising is the biggest application of the data that is collected about you at the moment, but there are other uses thinkable which may already be in use albeit in a less open way as advertising.

It will be very interesting for insurance companies to know about which kind of diseases you've been searching for online. This may tell them how big a risk you are for them and what the changes are that you may be knocking on their door for money. If they deem this change very high they may heighten the fee you have to pay them. In this case the data that is collected on the internet becomes another part of your file with these agencies determining whether you'll get cheap insurance or not. This may seem far fetched or unethical to you, but the fact is that insurance companies have done similar things in the past. In the Netherlands for instance insurance companies have payed a freelancer to talk tax officials in giving away social security numbers of people who the companies wanted to check. It wouldn't surprise me if some companies, under false pretenses, were already busy gathering your medical data and selling that to interested parties. There is a lot of money to earn with insurances!

Another application of this data and which is similar to insurance companies is with banks. When you go to a bank to get credit the bank will do everything it can to make sure you're good for your money. Seeing whether you gamble or tend to spend a lot in shops may be an indication for banks to trust you or not.

The third field where this kind of data can be useful is employment. An employer may want to know more about you and ask a company to generate a report about your online behavior. Before you can say anything about yourself an image about who you are is already made and such first impressions often count.

For all these uses of information counts that the companies asking for the information may not care about whether the data is giving a correct image. It's likely that it is more profitable for them to assume the data is correct and damage some people unjustly then take the risk of loosing a lot of money by trusting an entire group they'll know is likely to act against their own interest. You won't get benefit of the doubt and they won't try to talk about it if only because they don't want this sort of tracking

practice to be known by the public.

A completely different, but also profitable way of using your online behavior is what some travel agencies have been doing. They measure how often you would look at a certain flight. If you looked often or flew regularly they would heighten the prices for you only, because they make the guess that since you've been thinking about it for a while already or you regularly take that flight you are willing to pay a little more than others.

Of course it is not only companies who can have benefits from this sort of data. An employee from a perfectly behaving company may decide to sell data to interested companies for some extra money. Burglars may pick up on travel plans together with your address and people running political campaigns may be interested in what you think is important to present you a digital pamphlet that addresses all those issues you're worried about.

I'm sure that this long list of possible uses is limited and that there are uses people haven't even thought about. The worrying thing is that this kind of practice is usually well hidden until it goes wrong for the companies and a scandal emerges. Even if this way of doing business is revealed it doesn't always get enough media attention to change things.

Demystify language of ads

- Take apart inductive arguments made and criticize with help of Ian Hacking
- 1) Doubtful arguments made when you think ads become interesting once you clicked on them.
- 2) Problem with scale is that low percentages are profitable, but low percentages are not also user friendly.

Ways to intervene in the technology

Reasons why tracking is good

- marketing purposes
- machine learning
- sociological data
- api

Paradoxes between free/tracking and sharing/copyright

- what does it mean for things to be free?
- what does it mean for people to share?
- how may an intervention disrupt free and sharing?
- what could the economical damages?
- how much is an user willing to understand?

Argument against cumulation of capital

- The Mayfair Set by Adam Curtis
- Dark Side of Google (lack of innovation is happening)

Routing leaks?

By about a hundred professionals who occupy themselves with how to score high in Google it is estimated that the use of the website as monitored by Google will become more important. This means that pages which are used intensively will rank higher then they do today. For this mechanic to work it is necessary that the behavioral data is recorded.
http://www.seomoz.org/article/search-ranking-factors

Since 2003 Google has started to work with targeted advertisements. In their news item they claim that advertisers and partners get a high return on investment. The weird thing is that they also claim that users get a high return on investment. This bares the question what a user is actually investing and how users are rewarded for this investment.

Google seems to make the argument that if Google makes a lot of money, because there is a high click-through-rate the users must be rewarded for their investment of their privacy as well. This is an invalid argument, because the big profit can also derive from the fact that Google works on a huge scale. If only 10% of the people click through Google's profit must already be enormous. But it actually means that the percentage of users happy with targeted advertisements is quite low while they run the risk of a privacy breach.
http://www.google.com/press/pressrel/advertising.html

When Google started to combine information from different profiles there was reason for some to look for alternatives. There are a few things to think about when protecting your privacy:

- search engine
- email
- browser (search engines in browser)
- web beacon scripts

http://www.security.nl/artikel/40063/1/Een_leven_zonder_Google.html

Googles new privacy policy doesn't change much for Google Analytics. They still won't share visitors data with other parties or even other Google services. However they are busy with a program that seems to be an experiment to use Google Analytics data in the same way they use other gathered information.
http://analytics.blogspot.com/2012/01/googles-updated-privacy-policy-what-it.html

It's not very important to focus upon something like fingerprinting. I say this, because 97% of the data leaks that occur on the internet are caused by SQL-injection. SQL-injection is an outdated principle, but it seems that it is still in use vigorously. There are even people pleading to stop trying to secure other kind of leaks before SQL leaks are solved. The same can be said in regard to cookies and browser fingerprinting. Cookies are an old technology, but as long as they are mostly responsible for giving identity away I say we should focus upon them.
http://www.cio.co.uk/news/3331490/barclaycard-97-per-cent-of-data-breaches-due-sql-injection/

Google has launched a new service where you can search for symptoms and then Google will do suggestions about the kind of disease you have. Ads are not personalized based on this information and it's possible to wipe your search history, but it remains unclear if websites can pick up upon search terms.
http://www.readwriteweb.com/archives/why_google_didnt_build_search_plus_your_body.php

If Republican presidential candidates are mentioned by name on Facebook a company will automatically measure the mood of the message. This is seen as interesting sociological information by

some and a privacy nightmare by others. The biggest privacy objection is that there is no way to turn this feature off. An argument for this kind of action is that sociological data in the past has helped to fight discrimination.
http://www.readwriteweb.com/archives/why_facebooks_data_sharing_matters.php?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed%3A+readwriteweb+%28ReadWriteWeb%29

With frictionless sharing, Facebook has changed sharing into archiving peoples consumer habits. Everything you see and read (or actually … click on) will be put into your Facebook news feed. The author of the source has a strange double attitude towards it. I think it's wrong to see me as a consuming machine with billboard properties.
http://www.readwriteweb.com/archives/facebook_hasnt_ruined_sharing_its_just_re-defined_it.php

Google started a big “privacy” campaign about internet. However it doesn't seem to be addressing the issues I'm worrying about in relation to Google.
http://www.security.nl/nieuwsbrief/artikel/187/39931?utm_source=newsletter&utm_medium=email&utm_campaign=newsletter

It was shown by research that a lot of people do not understand privacy tools such as opt-out, Ad Block, Ghostery and others. They were less well protected than thought or weren't protected at all. Filters were the hardest to understand. The things people run into are familiar. It is a sign that there is still a lot of opportunity to improve privacy tools.
http://www.security.nl/nieuwsbrief/artikel/175/39042?utm_source=newsletter&utm_medium=email&utm_campaign=newsletter
http://www.sciencedaily.com/releases/2011/10/111031120249.htm

You can now opt-out for Google targeted advertisement.
http://www.security.nl/nieuwsbrief/artikel/175/39044?utm_source=newsletter&utm_medium=email&utm_campaign=newsletter

Google admits that its business model is in conflict with privacy, Encryption is no option since Google makes money by selling targeted advertisements and this is impossible if Google can't get to the data.
http://www.security.nl/nieuwsbrief/artikel/175/39067?utm_source=newsletter&utm_medium=email&utm_campaign=newsletter
http://www.businessinsider.com/googles-business-model-is-in-conflict-with-your-privacy-2011-11

Facebook is having trouble to make profit. Advertisements are being more ignored than anticipated. This is why Facebook will take a number of actions to make ads more invasive to the user experience.
http://www.privco.com/recently-leaked-facebook-documents-show-facebook-scrambling-for-improved-ad-performance-to-boost-short-term-revenues-for-ipo-company-trailing-revenue-plan
http://venturebeat.com/2012/02/23/facebook-q1-ad-revenue/

Users are expecting miracles from the Do Not Track functionality. In truth the companies that will take into account the header will not show targeted commercials, but any tracking could still happen in the background.
http://www.security.nl/nieuwsbrief/artikel/192/40505?utm_source=newsletter&utm_medium=email&utm_campaign=newsletter

The new privacy policy of Google especially hits Android users, because to use that device effectively

you'll need an account and everything you do can now be linked to this device. Canadian officials are concerned.
http://www.security.nl/nieuwsbrief/artikel/192/40503?utm_source=newsletter&utm_medium=email&utm_campaign=newsletter

Google in trouble because it circumvented default privacy settings of both Safari and Internet Explorer. Google followed them by mistake according to Google.
http://www.security.nl/nieuwsbrief/artikel/192/40426?utm_source=newsletter&utm_medium=email&utm_campaign=newsletter
http://www.security.nl/nieuwsbrief/artikel/192/40416?utm_source=newsletter&utm_medium=email&utm_campaign=newsletter

Google is paying people to join in two programs that are more intrusive than their current ways of tracking people online. The amount of data gathered is similar to what Google Analytics gathers and in one of the cases it is even more than that!
http://www.neowin.net/news/google-paying-people-to-track-their-web-visits
http://www.security.nl/artikel/40256/1/Google_betaalt_internetter_20_euro_voor_privacy.html

Chrome users are being followed to detect malicious phising sites and downloads. IP addresses are gathered, but are stripped from the URL after two weeks.
http://www.security.nl/artikel/40141/Google_gebruikt_Chrome-gebruikers_als_honeypot.html

According to some scientists it is very hard to de-identify people. Presumably this is because two data sets can be linked together and make the group of possible matches a lot smaller. However according to this article it is hard, but not impossible to de-identify. It is good to do this kind of thing, because it may help public health if data is looked at.
http://www.sciencedaily.com/releases/2011/06/110616092650.htm

Twitter was uploading user contacts without informing them. They admitted this and you can now remove the data, the question remains how many people know about this? Another startup called Path did a similar thing.
http://www.readwriteweb.com/archives/twitter_is_the_latest_company_to_admit_it_uploads.php

A burglar was using Google Maps to scan the area before he striked. It's possible to remove your house from Google Maps.
http://www.businessinsider.com/how-to-protect-your-home-from-a-google-maps-burglar-2011-9

The new privacy policy of Google is there to allow Gmail related data to be used in Youtube and vice versa. This is probably done to get more income from advertisement.
http://www.businessinsider.com/heres-the-crucial-part-of-googles-new-privacy-policy-that-has-advertisers-salivating-2012-1

Googles side of the story (concerning 3rd party cookies in Safari) is that people opted-in for +1 buttons on advertisements. Somehow Doubleclick also got access through the loophole made for these +1 buttons.
What we can also conclude from this is that Safari doesn't block anything as long as users once trusted the sites from trackers. This makes from a moderate privacy protection I would say.
http://www.businessinsider.com/google-apple-tracking-explanation-2012-2

When you use an Android phone and ads appear in your browser. Your phone number is connected to the behavioral data according to this source. However the source doesn't think this is a risk in anyway and again the argument for beneficial advertisements show up.
http://www.businessinsider.com/why-you-want-google-to-know-everything-about-you-2012-2

A science fiction story at the moment, but it may be that salespersons or people giving away flyers will be prompted to reach you based on your Google settings.
http://www.businessinsider.com/how-google-apps-may-soon-let-salespeople-stalk-you-2012-2

A very violent text about how privacy doesn't matter anymore.
http://www.businessinsider.com/online-privacy-who-cares-2012-2

A very good text about how privacy does matter, but is a complicated whole. It tells a little about the companies that are following you!!!
http://www.theatlantic.com/technology/archive/2012/02/im-being-followed-how-google-and-104-other-companies-are-tracking-me-on-the-web/253758/

More official source about how web beacons work.
http://priv3.icsi.berkeley.edu/

By leaking a large amount of personal data it's possible that users become victim of identity theft. Loosing this data also helps to de-anonimize other databases. It has been shown that such data can do that.
http://www.sciencedaily.com/releases/2012/01/120118122829.htm

It may be a good idea to split up different data sets about oneself for different purposes.
http://www.sciencedaily.com/releases/2008/12/081204094555.htm

75% of websites leak data to third parties. 50% of the websites leaks identifying numbers, which can be cross matched to identify somebody. It's impossible to say which companies are bad and which aren't. In the mean time it is clear that there are high incentives to collect data and that those companies can't be trusted to protect privacy. The responsibility should lie with 1st parties according to the article. All the consumer tools available leave some kind of hole in the security according to this report.
http://www.sciencedaily.com/releases/2011/06/110602111437.htm

Data about Google employees was stolen (through burglary) which can be used to perform identity theft and make credit card accounts on behalf of others.
http://news.cnet.com/Stolen-Google-employees-personal-data/2100-1029_3-6243093.html

Both Facebook and Google have fired people, because employees were accessing personal data from users that they shouldn't look into. According to Google it has only happened twice in 10 years time with 20.000 people working there. A very low number which may be false, but Facebook isn't commenting at all. Apparently somebody who was working at the help desk also misused his knowledge power and harassed teens.
http://techcrunch.com/2010/09/14/google-engineer-fired-security/
http://thenextweb.com/google/2010/09/14/ex-google-employee-dug-through-private-data-and-harassed-teens/

If you searched for herpes, then the search term together with your IP address and possibly a cookie

will be recorded. This could lead to unwanted things with acquiring credit and health insurances.
http://www.security.nl/nieuwsbrief/artikel/187/39911?utm_source=newsletter&utm_medium=email&utm_campaign=newsletter